2

DEPARTMENT
of
PSYCHOLOGY

Carnegie Mellon University

[redacted]

(2)

Final Report on ONR Grant N00014-88-J-1153

Architectures for Intelligence: The Twenty-Second Carnegie Symposium on Cognition

October 5, 1990

Kurt VanLehn
Learning Research and Development Center
University of Pittsburgh

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; Distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Carnegie 1	5. MONITORING ORGANIZATION REPORT NUMBER(S) Same as Performing Organization
6a. NAME OF PERFORMING ORGANIZATION Carnegie Mellon University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Computer Science Division Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) Department of Psychology Pittsburgh, Pennsylvania 15213	7b. ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, VA 22217-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Same as Monitoring Organization 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-88-J-1153
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. N/A · PROJECT NO. N/A · TASK NO. N/A · WORK UNIT ACCESSION NO N/A

11. TITLE (Include Security Classification) Architectures for Intelligence: The Twenty-Second Carnegie Symposium on Cognition (Unclassified)

12. PERSONAL AUTHOR(S) Kurt VanLehn

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 88May1 TO 89/4/30	90/10/12	29

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Artificial Intelligence, Cognitive Architectures Cognitive Psychology, Cognitive Science,

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This proposal requests funds to partially support a symposium on Architectures for Intelligence. The major purposes of the symposium are: (1) To promote interaction among researchers who are pursuing the architectural question from divergent viewpoints. (2) To exhibit the common issues in architecture research that may have been obscured by the variety of approaches. (3) To see if there are a common set of "good ideas" that crop up in a variety of architectures. (4) To compare varying degrees of ontological commitment, which range from "an architecture is just a notation for computations, and any convenient one would do as well" to "there is one optimal architecture, both for the human mind and the artificial mind, and our architectures are hypotheses about what that real architecture is." (5) To examine the "levels of description" idea, which is used, for instance, to say that connectionist architectures describe the mind at a finer-grained level of description than serial, symbolic architectures, so both descriptions can be right at the same time. The speakers at the symposium have all written chapters for a book entitled *Architectures for Intelligence*, which is being published by Erlbaum and should appear in March, 1991. Keywords:

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED XXX SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL Susan Chipman	22b. TELEPHONE (Include Area Code) (202) 696-4322 · 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

Abstract

This proposal requests funds to partially support a symposium on Architectures for Intelligence. The major purposes of the symposium are: (1) To promote interaction among researchers who are pursuing the architectural question from divergent viewpoints. (2) To exhibit the common issues in architecture research that may have been obscured by the variety of approaches. (3) To see if there are a common set of "good ideas" that crop up in a variety of architectures. (4) To compare varying degrees of ontological commitment, which range from "an architecture is just a notation for computations, and any convenient one would do as well" to "there is one optimal architecture, both for the human mind and the artificial mind, and our architectures are hypotheses about what that real architecture is." (5) To examine the "levels of description" idea, which is used, for instance, to say that connectionist architectures describe the mind at a finer-grained level of description than serial, symbolic architectures, so both descriptions can be right at the same time. The speakers at the symposium have all written chapters for a book entitled *Architectures for Intelligence*, which is being published by Erlbaum and should appear in March, 1991.

Final Report on ONR Grant N00014-88-J-1153

Architectures for Intelligence: The Twenty-Second Carnegie Symposium on Cognition

Kurt VanLehn
Learning Research and Development Center
University of Pittsburgh

October 5, 1990

Executive Overview

This grant partially supported a symposium on *Architectures for Intelligence*. The Symposium, was held on May 16 and 17, 1988. It was the twenty-second in a series of Carnegie Symposia on Cognition. The papers presented at the Symposium are being published by Erlbaum. The presenters came from a variety of disciplines, including cognitive psychology, knowledge engineering and robotics. They all have developed computing architectures that run programs that exhibit intelligence of various kinds. Some presenters view their architectures *descriptively*, as models of the human mind. Other presenters view their architectures *prescriptively*, as infrastructures for development of optimally intelligent machines. The descriptive and prescriptive views are two sides of the same coin. They are simply two different methods for coming to understand the architectural principles that underlie intelligent information processing.

The major purposes of the symposium were: (1) To promote interaction among researchers who are pursuing the architectural question from divergent viewpoints. (2) To exhibit the common issues in architecture research that may have been obscured by the variety of approaches. (3) To see if there are a common set of "good ideas" that crop up in a variety of architectures. To compare varying degrees of ontological commitment, which range from "an architecture is just a notation for computations, and any convenient one would do as well" to "there is one optimal architecture, both for the human mind and the artificial mind, and our architectures are hypotheses about what that real architecture is." (4) To examine the "levels of description" idea, which is used, for instance, to say that connectionist architectures describe the mind at a finer-grained level of description than serial, symbolic architectures, so both descriptions can be right at the same time.

Background

Thoroughout the history of research on architectures for intelligence, which goes back to the mid-1950s, there have been seemingly separate, parallel developments in cognitive psychology and artificial intelligence. However, the same issues have arisen in both literatures, often at about the same time, although sometimes under different names. For instance, the earliest architectures had a serial processor with a small amount of temporary working memory. The psychologists called the working memory "STM" and the computer scientists called it "registers."

In the late-1960's, both cognitive psychology and AI attacked the issue of control. Both were concerned that early types of control regimes (the AI term) and attention mechanisms (the psychology term) were too rigid and could not react flexibly to changes in the environment. Later, much research was focused on the organization of factual knowledge. During this period, the literatures of AI and cognitive psychology began to merge. Semantic nets, schemas and frames became common both in psychological work and AI work.

In the late 1970's, perhaps the major architectural issue became practice effects. An obvious way that people adapt (and machines should adapt) to finding themselves in a repetitive lifestyle is to automatically acquire speed, accuracy and automaticity through mere repetition of an activity. This led a variety of proposed architectural mechanisms, including macro-operator formation, chunking, production compounding, and explanation-based learning.

Currently, much attention is being focused on transfer. People have the ability (and machines should have the ability) to go beyond their training and successfully blend elements of the knowledge that they have acquired to solve novel problems. Some architectures, such as the connectionist networks, place this blending deep in the basic functioning of the architecture. Others place it higher, as a weak method of problem solving, called problem solving by analogy. Midway between these extremes are mechanisms such as production overlap, wherein transfer occurs via a mixture of shared knowledge structures and strategy.

Another current issue is the ability of the human mind to reflect on its own processing. Although this ability has been deliberately ignored by American psychology since the behaviorist's scathing attacks on introspective methods, there is no doubt of its existence. This raises some interesting questions, such as how can the mind reflect and what is reflection good for? What is the relationship between a person's plans and their (situated) actions? So far, research on these questions has been done mostly by AI researchers and anthropologists; perhaps it is time to bring that work to the attention of cognitive psychologists.

A third current issue, one with an especially long history in computer science, is parallel processing. Organizing parallel processing to take full advantage parallelism remains an unsolved problem, although good results have been acheived for narrow classes of problems. Yet, the neurological evidence seems to show that the human brain is a massively parallel computer. The issue is to find some task-general (e.g., architectural) way of organizing parallel computing. Connectionism has made dramatic inroads on this problem, and the fruits of that exploration are now being abstracted and applied in non-connectionist architectures.

However, the parallelism problem, and in fact, all the architectural problems mentioned above, are still being actively researched. While research on architectures tends, like all research areas, to wax and wan over the years, interests in architectures is currently on the rise. Thus, "Architectures for Intelligence" is a particularly fitting topic a Carnegie Symposium.

Moreover, throughout the history of research on architectures, the Carnegie Symposia have been an important cross-pollination site for research efforts from disparate fields. The 1988 symposium continued that tradition by attempting to hybridize architectural research from AI and cognitive psychology.

The presentations

The symposium was organized as ten presentations with two commentaries. Prof. Herbert Simon later contributed a third commentary for the book, although he did not speak at the Symposium. This section describes the 13 contributions. The Symposium and the book are divided into two parts. The first part includes works that are primarily descriptive -- they undertake to accurately characterize some aspect of the human mind. The second part includes works that are primarily prescriptive -- they propose and evaluate designs for intelligent artificial architectures. Each part concludes with a commentary on the presentations in that part.

Part 1: Descriptions of the human architecture

John R. Anderson (Carnegie-Mellon University) *The status of cognitive architectures in a rational analysis.* Prof. Anderson argues that human cognition can be predicted from the assumption that it is optimized to the information-processing demands that are placed on it. He shows that results that are taken in support of particular architectures (PDP, ACT*, SOAR) are are consequences of this rationality principle of human cognition. Implications of this rationality principle for cognitive architecture were discussed.

Herbert A. Simon (Carnegie-Mellon University) *Cognitive architectures and rational analysis: Comment.* Prof. Simon begins by reviewing arguments from economics and psychology that attempt to explain phenomena as resulting from optimal adaption to the environment. He demonstrates that there are plausible alternative explanations for the phenomena that Anderson takes as support for his position. Simon concludes by reviewing his well-known position on bounded rationality.

James L. McClelland and **Eric Jenkins** (Carnegie-Mellon University) *Nature, nurture and connections: Implicatiosn of connectionist models for cognitive architectures.* Prof. McClelland and Mr. Jenkins argue that connectionist architectures form the bais for the acquisition of a number of cognitive abilities. They demonstrate the explanatory power of their hypothesis with three well-known developmental phenomena: (1) failures of conservation and compensation; (2) progressive differentiation of knowledge about different kinds of things, and (3) U-shaped learning curves in language acquisition.

Paul S. Rosenbloom (Information Sciences Institute), **Allen Newell** (Carnegie-Mellon University) and **John E. Laird** (University of Michigan) *Towards the knowledge level in Soar: The role of the architecture in the use of knowledge.* Soar has been described as an architecture for a system that is to be capable of general intelligence. One way to specify what this might mean is to define general intelligence as the ability to approximate an ideal knowledge level system across a sufficiently broad set of goals and knowledge. In this chapter we use this definition as the basis for evaluating the degree to which Soar achieves general intelligence. A complete evaluation is beyond the scope of this chapter, so we focus more narrowly on how the Soar architecture supports and constrains the representation, storage, retrieval use and acquisition of three pervasive forms of knowledge: procedural, episodic and declaritive knowledge. The analysis reveals that Soar adequately supports procedural knowledge -- to some extent it was designed for this -- but that there are still significant questions about episodic and declarative knowledge. These questions arise primarily because of consequences of the principle source of constraint in Soar, the fact that all learning occurs via chunking. New results are also presented on teh acquisition of declarative knowledge.

Walter Schneider and **William L. Oliver** (University of Pittsburgh) *An instructable connectionist/control architecture: Using rule-based instructions to accomplish connectionist learning in a human time scale.* Schneider and Oliver have been studying the time course of human skill acquisition. They have found that standard connectionist learning seems unable to acquire skills within the lifetime of a human learner. They have constructed an architecture, Cap2, that is a hybrid of a connection system and a symbol-processing system. Cap2 is an instructable connectionist system whose learning performance fits the observed learning performance of humans with greater fidelity than its antecendents. Cap2 embeds a variety of novel assumptions about traditional architectural issues, including working memory, control structure, knowledge organization, learning during practice, transfer, parallelism and even reflection.

Kurt VanLehn and **William Ball** (University of Pittsburgh) *Goal reconstruction: How Teton blends situated action and planned action.* Prof. VanLehn has been studying how people utilize routine procedures, such as the well-learned problem-solving schemas of experts. He has found that even the most routine plans are often not followed in a simple, straightforward fashion. Other researchers in cognitive science, notably ethnomethodologists and psycholinguists, have also found that the relationship between the a person's supposed plans and their actual actions is much more complex than had ever been imagined. However, instead of rejecting plans as epiphenomenal, VanLehn argues that the complex relationship between plans and actions can in fact be understood, and indeed, appears to fall out as a side-effect of limitations of working memory. He has constructed an architecture, Teton, that explicates the relationship between plans, actions and working memory.

Zenon Pylyshyn (University of Western Ontario) *The role of cognitive architectures in theories of cognition.* Prof. Pylyshyn begins with an clarification of the posible roles that cognitive architectures could have in explanatory accounts of human cognition. With this background, he comments on some of the specific points raised by speakers in part 1 of the symposium.

Part 2: Prescriptions for artificial architectures

Rodney A. Brooks (Massachusetts Institute of Technology) *How to build creatures rather than isolated cognitive imitators.* Prof. Brooks has been developing an architecture for mobile robots called the subsumption architecture. The key idea is that the unit of modularization of the subsumption architecture is a task, such as avoiding obstacles, rather than a function, such as perception. Thus, the avoid-obstacles task-module has specialized perception, cognition and locomotation functions integrated into it. All task-modules run in parallel, roughly speaking, and the subsumption architecture (which is more of a design philosophy than an architecture, as it is standardly conceived) handles their interactions. Brooks argues that the subsumption architecture is currently adequate for modelling lower forms of animal life, such as insects, and that higher forms of intelligence can be obtained by the accretion of more and more task-modules (the correspondence between such accretion and evolution are obvious, exciting and highly controversial). Although Brooks' approach represents a fairly radical break from traditional architectures, it nonetheless bears strongly on some of the architectural issues that have arisen over the years. In particular, it bears directly on the transfer issue and the parallelism issue. Brooks' group has build a series of complete creatures (Allen, Herbert, Tom and Jerry, and now Seymour under construction) which exist in and interact with ordinary people-populated office and laboratory areas, and thus must go about their business in an unstructred dynamically chaning environment.

Jamie G. Carbonell, Craig A. Knoblock (Carnegie-Mellon University) and **Steve Minton** (Nasa Ames Research Center) *Prodigy: An integrated architecture for planning and learning.* Prodigy is a computational architecture that integrates general problem solving and multiple learning methods. The primary design objectives are: to provide an open-architecture research vehicle to gain insight into deliberative symbolic reasoning, to investigate learning in the context of a performance engine, to permit the evaaluation of multiple learning techniques within the same architecture in multiple domains, and to provied a basis for the development of flexible, adaptive knolwedge-based systems. The Prodigy system consists of a general planning and problem-solving engine, a set of machine learning techniques, and multiple knowledge sources encoded in a uniform, logic-based knowledge representation. In particular, Prodigy learns control knowledge via explanation-based learning and static domain analysis, forms abstraction hierarchies for effective planning, recyles past experience via derivational analogy, extends and refines domain knowledge through experimentation, and acquires knowledge dynamically from domain experts. Prodigy has been tested in varioius domains such as basic machine-shop scheduling and high-level robitic planning. THis paper focuses primarily on the general Prodigy problem solver, the explanation-based learning method, the abstraction learning method, and an empirical evaluation of these methods on large populations of problems.

Michael R. Genesereth (Stanford University) *A comparative analysis of some simple architectures for autonomous agents.* Prof. Genesereth undertakes a comparative analysis of several simple architecures, some of which use declarative knowledge and some of which do not. He shows that there is indeed no advantage to declarative knowledge at runtime if we have unlimited space and ignore the costs of agent design. However, he also shows that, in the face of space limitations or considering design costs, the runtime use of declarative knowledge is sometimes superior to preprogramming.

Barbara Hayes-Roth (Stanford University) *Making intelligent systems adaptive.* Prof. Hayes-Roth has been tackling the problem of integrating perception, action and cognition, and doing so in a way that allows adaption of the system to the demands of a dynamic environment. A central issue is the real-time control of attention. As a testbed for her design principles, Hayes-Roth has been developing Guardian, a system that monitors patients in a surgical intensive care unit.

Tom M. Mitchell, John Allen, Prasad Chalasani, John Cheng, Oren Etzioni, Marc Ringuette and **Jeffery C. Schlimmer** (Carnegie-Mellon University) *Theo: A framework for self-improving systems.* Theo is a software framework to support development of self-modifying problem-solving systems. It provides a uniform representation in which beliefs are represented as values of slots of frames, and problems are represented by slot instances whose values are not yet known. Theo can also represent meta-beliefs and pose meta-problems about any of its beliefs or problems, including its meta-beliefs and meta-problems. This representation allows it to describe information about problem-solvin methods and strategies in terms of explicit beliefs about the problems for which these methods are appropriate. In addition, the meta-beliefs are indexed by the ground beliefs so that their retrieval const does not grow significantly as the size of the knowledge base increases. This chapter discusses the motivation and goals underlyhing the design of Theo, and provides an overview of the present incarnation of the system. Experimental results are presented showing the impact of three learning mchanisms currently implemented in Theo: caching of inferred beliefs, explanation-based learing of macor-methods for problem solving, and inductive inference of control information for guiding problem solving. The authors also discuss the relationship between Theo and two earlier frameworks with related goals: Soar and RLL.

William J. Clancey (Institute for Research on Learning) *The frame of reference problem in the design of intelligent machines: Commentary on the twenty-second Carnegie Symposium on Cognition.* Is knowledge a substance that can be stored? When people reinterpret what the symbols in a knowledge base mean, does that change what the program knows? How is a deisigner to construct an intelligent machine without bounding its behavior by his own preconceptions about the world? Such questions are analyzed here interms of "frames of reference" in the modeling of complex systems. A framework is provided that distinguishes between a robot designer's ontological preconceptions, the dynamics of a robot's interaction with an environment, and an observer's descriptive theories of patterns in the robot's behavior. Claney argues that knowledge-level descriptions are fundamentally about an interactive, social system, not individual agents; that they consitute an observer's theory, not representations possessed by the subjects being studied; and they characterize emergent patterns, not plans necessarily precast by the agent. From this perspective, the relation of knowledge-level descriptions to the functional architecture (physical mechanisms) of intelligence is reconsidered, integrating the work of Newell, Pylyshyn and Dennett. A functional architecture that manipulates processes rather than descriptions or labelled structures is proposed, and the objectives and approaches of papers from the 22nd Carnegie Symposium on Cogntion are re-examined.

Current Status

The speakers at the symposium have written chapters for a book entitled *Architectures for Intelligence.* The book is being published by Erlbaum as part of their Carnegie Symposium Series. At the present time, all the chapters have been copy-edited and all but three authors have checked the copy-editing and returned their manuscripts to Erlbaum. When the remaining chapters have been returned, typesetting, proofchecking, indexing and binding are estimated to take 5 months. This means the volume should appear in March, 1991.